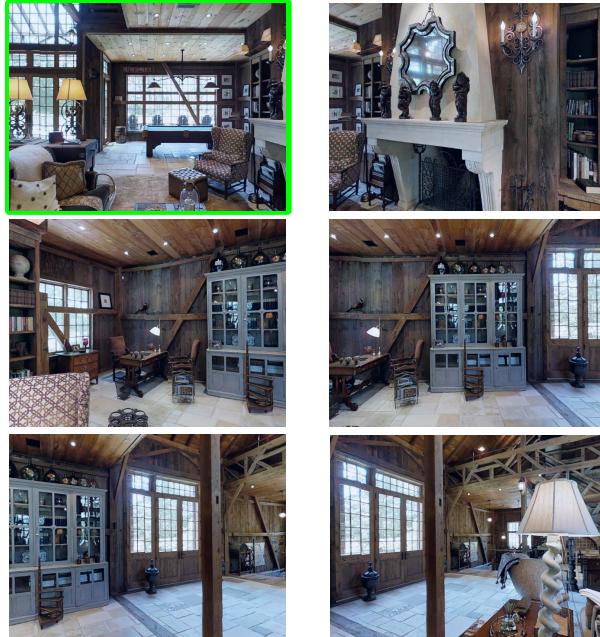


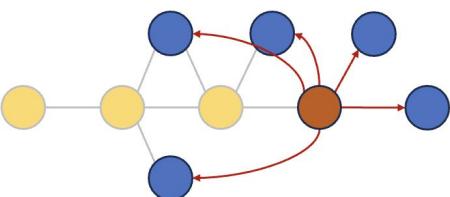
Original Instruction

"Exit the stairs, and then make a right. Walk nearby a sofa and then make a right. Continue walking until you reach game room with a pool table and then wait there."

Visual Observations



Topological Map



Reordered Instruction

1. Exit the stairs.
2. Turn right.
3. Walk nearby a sofa.
4. Turn right.
5. Walk to the game room.
6. Wait at the pool table.

Temporal Reordering Module



Subgoal

5. *Walk to the game room.*

VLM-based Action Router



Top-1 Routing

Prompt for Reordering

Turn navigation instructions into clear, **step-by-step actions** for an agent by breaking them into short, goal-focused steps. Make all **hidden temporal or spatial cues explicit** and preserve the **correct order** of actions.

Prompt for Router

Phase 1. Subgoal Localizer: Identify the next navigation step from the reordered instruction, using visual history and prior completed subgoals.

Phase 2. Skill Router: Classify the **primary skill** needed to execute the selected sub-instruction with the **full instruction, subgoal** to be executed (from Phase 1), and **reasoning** behind the sub-goal selection.

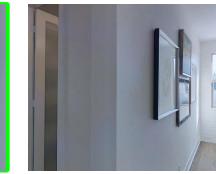
Skill-based Agents



Direction Adjustment



Take a right turn.



Vertical Movement



Go to the upper level.



Landmark Detection



Move past the bed.



Enter the kitchen.



Stop and Pause



Stop by the wheelchair.



Action Image



- Fine-tuned Model
- Inference Model
- Activated Expert